International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-01, Apr 2018

Survey on Machine Learning Techniques Used for Stock Market Prediction

[1]Saravanan Ramalingam, [2]Dr. Sujatha Pothula

[1]Research Scholar, [2]Assistant Professor, [1,2]Department of Computer Science, Pondicherry University, Puducherry, India.

[1]rsaravanan26@gmail.com, [2]spothula@gmail.com

***Abstract*—Stock market prediction is a wide area of research where a lot of prediction models and algorithms have been proposed focusing on predicting the stock market movements or the stock price changes of a particular company or a given set of companies using fundamental or technical analysis. It can significantly help investor in his decision making by recommending him/her the stocks that are probable to perform well. Among the various algorithms, due to the chaotic nature of the stocks and the randomness of the market, the machine learning algorithms proved to be the most popular for predicting the stocks. In this work, a comparison of the various machine learning algorithms for Stock market prediction such as Support Vector Machines, Naïve Bayes, Random Forest, K-nearest neighbours, Artificial Neural Networks, Regressions, Radial Basis Functions and other hybrid models is done based on the accuracy and various performance.**

Keywords— ***Stock Price Prediction, Machine Learning, Regression, ANN, Predictive Analytics***

I. INTRODUCTION

Stock market is one of the many ways people use to invest their hard-earned money. But investing in the Stock market has a high-risk factor. To aid the investors, Stock market prediction is being done. People gain money if the stock price rise or lose it if they go down. Around 90% of the people who trade in stock market lose their money due to various reasons.

According to "The Financial Express" [1], the major reasons for investors failing in stock market are

- Investing in the market without any fundamental knowledge of the market
- Investing based on advice from people who themselves rely on the advice of others
- Speculating the market situations and investing
- Lack of patience at times of crisis
- Having portfolios that are not professional
- Trying to follow others while investing
- Having unrealistic expectations of getting high profits in short time
- Lack of immediate action when one has to take extreme steps
- Doing no research on the company's performance before investing in its stocks
- Due to emotional conditions making changes in investment without thinking about the actual initial plan of investment

Only 5% of traders have proper knowledge of the share market and make use of appropriate prediction methods to gain in the market.

Stock market predictions are done so that the investment does not have a risk of loss in the monetary value. Stocks that give high percentage of profits or those which may lead to losses can be found by predicting the stock market, not only helping people to refrain from the loss that they may incur but also to gain considerable profits. To aid the decision making, a lot of methods have been proposed to predict the stocks for both short term as well as long term investments [2]. Short term investments in general are done by the individual persons whereas long-term investments are done by the professional traders and by organizations.

Stock Market predictions are done using various methods viz. Technical Analysis, Time series forecasting, Machine learning and data mining and Modelling and predicting volatility of stocks using Differential equations [3]. Among these methods, Machine Learning and Data mining is widely used prediction method for stocks due to the chaotic nature of the stock market. Moreover, the large amount of stock data has a lot of noise from where a valid data should be found which gives useful information by identifying the patterns in the data and help in predicting the stocks. A large number of machine learning algorithms have been proposed for predicting the stocks such as Support Vector Machines, Naïve Bayes, Regressions, Radial Basis Functions, Random Forest, k-Nearest Neighbor, Artificial Neural Networks and many more. In this work we present a detailed comparison of the performance analysis of the stock prediction using the above said algorithms.

II. FEATURES USED FOR STOCK PREDICTION

Stock Price Prediction is carried out with the help of the large historical datasets that are collected for the particular stock under prediction. These datasets consist of a set of attributes, from which various technical indicators that are used for prediction are derived. The effectiveness of the various models is traced using various performance factors. The following lists few of those attributes, technical indicators and performance factors that are being used for prediction.

 DOI : 10.18231/2454-9150.2018.0109

A. Attributes used in Stock Market

The following are the primary attributes of stocks used in stock market prediction.

- Company : Name of the company whose stocks are traded
- Stock : Name of the stock market in which the trade is taking place
- Date : The date of the day of trade
- Open Price : The price of the stock at the open time of the trading period
- High Price : The highest price of the stock during the trading period
- Low Price : The lowest price of the stock during the trading period
- Close Price : The price of the stock at the close time of the trading period
- Adjusted Close Price : The close price of the stock that is due to the trades in the other markets while the current market is closed
- Volume Traded : The amount of stocks traded over the trading period

B. Indicators of Stock Market

Stock Market uses two kinds of Indicators which includes Fundamental Indicators and Technical Indicators. The following is a list of few of the Technical Indicators used to predict the movement of stock indices.

Relative Strength Index (RSI):

RSI is a momentum indicator used to find the speed and change in the price movements.

$$100 - \frac{100}{1 + (\sum_{i=1}^{0} Up_{t-i} / n) / (\sum_{i=0}^{n-1} Dw_{t-i} / n)} \quad (1)$$

Stochastic Oscillator:

It is a momentum indicator used to compare the security close price to a range of prices over a given time period.

$$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} * 100 \quad (2)$$

Williams %R:

It is a momentum indicator used to indicate whether a stock is overbought or oversold

$$\frac{H_n - C_t}{H_n - L_n} * 100 \quad (3)$$

Moving Average Convergence/ Divergence (MVCD):

It is an indicator used to find the momentum and directional strength of a particular stock

$$\text{OSC(t)-EMAosc(t)} \quad (4)$$

Rate of Change (ROC):

It is the change in close price over the past n days

$$\frac{C_t}{C_{t-n}} * 100$$

Momentum:

It is a measure of the change in security close price over a period of time

$$C_t - C_{t-4}$$

Commodity Channel Index (CCI):

It is a measure of the variation of security close price to its statistical mean

$$\frac{M_t - SM_t}{0.015 D_t}$$

Average Directional Index (ADI):

It is used to discover trends in the data that is developing over time

$$\text{Sum((+DI-(-DI))/(+DI+(-DI))/n}$$

C. Performance factors used in stocks

Performance factors indicate how well the algorithm used for stock prediction works. The following are few performance factors used in the analysis of the algorithms in predicting the stock market.

Root Mean Square Error (RMSE):

It is a measure of the differences between values predicted by the model and the values observed.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y' - y)^2}$$

Mean Absolute Error (MAE):

It is a measure of the differences between two continuous variables.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y - y'\right|$$

Mean Absolute Percentage Error (MAPE):

It is a measure of the prediction accuracy of a forecasting method in statistical method.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y - y'}{y}\right|$$

Mean Square Error (MSE):

It is used to measure the average of the squares of the errors or deviations of the estimated and actual value

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y - y')^2$$

By using the various parameters listed above, a number of models have evolved over time. In the succeeding part of the work we present an overview of a few of those models that are used in Stock Market Prediction

III. Overview Of Prediction Models For Stock Market

Prediction in the stock market makes it easy for the buyers as well as sellers t (5) t properly. This is generally done using different prediction algorithms by different traders. In this paper, an overview of various machine learning algorithms used f tock market prediction has (6)

 DOI : 10.18231/2454-9150.2018.0109

been surveyed and compared based on their performance metrics.

Jhi-Sheng Chou et al, (2017) [2] use a combination Least Squares Support Vector Regression algorithm with firefly algorithm (MetaFA-LSSVR) to predict the stocks of the top five companies in the Taiwan Stock Exchange. The LS algorithm's efficiency depends largely on the hyper-parameters which are used for time series prediction. These hyper-parameters are tuned by the firefly algorithm in the MetaFA-LSSVR model. Here the hyper-parameters used include chaotic map, adaptive inertia weight and Levy flight. The training and testing dataset is separated using a sliding window mechanism. The performance of the algorithm is calculated using the values of RMSE, MAE, MAPE, MSE, correlation coefficient, non-linear regression multiple correlation coefficient and Synthesis index. The model provided a MAPE of 5.628 for 60 day prediction while it provided a MAPE of 6.131 for 90 day prediction resulting in predicting the short term investments better than the long term investments.

Luckyson Khaidem et al, (2016) [3] forecasted the direction of the stock market for Apple, Samsung and Microsoft using ensemble learning. The focus was to find the stocks with minimal risk of investment. The data was initially subjected to exponential smoothing and then predicted using Random forest (RF) algorithm. The technical indicators used were RSI, Stochastic Oscillator, Williams %R, MVCD. The Volume of stock trade was used to find the trends in the stock movement. The error in prediction was measured using Out Of Bag estimates. An accuracy of up to 80% for short term predictions and 85% for long term predictions were achieved.

Oscar Bustos et al, (2017) [4] compared the prediction of Colombian stock market using Artificial Neural Networks (ANN) and Support Vector Machine (SVM) with nine technical indicators. They predicted the stock price movements using technical indicators and performed automatic parameter tuning and cross validation to improve the performance of the models. The stock prices of around 25 companies under the Colombian stock market were predicted using both the models and an average of 76% for ANN and 78% for SVM was observed. But when compared with random test set ANN showed a high deviation of 20% using cross validation as opposed to SVM which had only a 2% change in value. The deviation observed was due to the generalization in ANN as compared to SVM.

Mahajan Shubhrata et al, (2016) [5] predicted the stock price by using daily checkout websites and view the information about share. Automation and prediction are used to improve the performance of system. Prediction is done in 2 ways - Dummy and Real time prediction. In dummy prediction model rules are defined to predict the future values and In Real time prediction the current share price of companies are predicted. Here the details such as name of the company, share and price are recorded. Naïve Bayes (NB) algorithm is used as classification technique for the given datasets. Dummy prediction and Real time prediction using NB algorithm provides accuracy of 75.6% to 76.5 %.

Xinjie Di, (2014) [6] predicted the trend in the Stock market using various Technical Indicators such as Williams %R, Rate of Change, Momentum, RSI, CCI, ADI, Triple Exponential Moving Average (TEMA), MACD, On Balance Volume, Time Series Forecasting, Average True Range and Money Flow Index. The algorithm used was Radial Basis Function (RBF) kernalized SVM and it was used for short term prediction of the stocks. After training with a data set of 1000 days it gave an accuracy of 56% for the Initial days which improved to 71% when stocks predicted for 10 days or after.

Bhagwant Chauhan et al, (2014) [7] viewed prediction of stock price level as a difficult monetary statistical problem. ANN is a mathematical model capable of machine learning and pattern matching. Artificial Neuron data or information is distributed through network is stored in the form of weighted interconnection. ANN has potential to uncover the unknown and hidden patterns in information which will be effective for share market prediction. Back Propagation (BP) Network, Multilayer forward Network learns by minimum sequence error and is employed in the area of language integration and adaption management. If the result yielded by the network is wrong, then the network scales back the chance of creating identical mistakes by learning. Data processing technique checks historical information of each day's open, high, low, close and adjusted close prices to desire value to improve accuracy. Accuracy of the algorithm is above 75%.

Haoming Li et al, (2014) [8] considering everyday open, close, high, low prices and trading volume tried to find the relation between the financial data collected and the actual performance of stocks using the linear models of Linear and Logistic Regression and SVMs. The data was collected from Quandl, Free online DB. Along with the financial data, 8 additional factors that include 5 commodity factors (Gold, Crude Oil, Natural Gas, Corn and Cotton), 2 foreign currencies (EUR, JPY) and 1 interest rate (10 year treasury rate) were considered. The relation of the stocks along with all the factors was calculated. Then from the various measured values four Direct Factors and one Indirect Factor was identified which was used to predict the direction of stock. with an accuracy of 56.65% with this model.

Khalid Alkhatib et al, (2013) [9] predicted the stock price of five major companies in the Jordanian stock exchange using k-Nearest Neighbour (kNN) Algorithm. Considering three attributes- low, high and close price and using the kNN classifier the classification is done using the similarities in the training dataset collected and the unknown new record. The 5-nearest records are found and the class is assigned using majority votes. The prediction values and errors are measured using Root Mean Squared Deviation, Explained Sum of Squares, and Average Estimated Error, Total sum of Squared Error, Average Error, Cumulative Closing price values. The deviation in prediction of the calculated values was up to 33% measured using the RMS values.

Yuqing Dai et al, (2013) [10] used various machine learning methods to forecast the stock price trend such as SVM, Logistic Regression, Quadratic Discriminant Analysis (QDA) and Gaussian Discriminant Analysis (GDA) methods. The data was collected from Bloomberg Data Terminal and 16 derived features were used for learning and developing the model. Some of the features are namely PE ratio, PX volume, current enterprise value, net price change, volatility, moving averages, quick ratios, risk premium, ISEPS and corresponding S&P-500 indices. Two

 DOI : 10.18231/2454-9150.2018.0109

models were constructed – Next day model and Long term model. The accuracy of prediction done is 58.2% for QDA in next day model and 79.3% for QDA in long term model which was better than the others

B.K. Panigrahi et al, (2012) [11] uses RBF which is an ANN function. It uses seven technical indicators such as Exponential Moving Average, Accumulation/Distribution Average, Stochastic Indicator, RSI, Price Rate Change, Close Price Acceleration and High Price Acceleration and five attributes, Open, High, Low, Close Prices and the volume of stock traded for prediction. The model is trained with a training set that has around 2510 patterns and is tested against FLANN and Multilayer Perceptron for its performance with MAPE. The comparison of the MAPE values show that the model performs better than the other two for both the DJIA and S&P500 stock Indices even though the deviation in prediction increases with time.

A Summary of the various stock prediction models reviewed is given in Table I.

TABLE I - REVIEW OF VARIOUS MACHINE LEARNING STOCK PREDICTION ALGORITHMS

Year	Author	Name of the Algorithm	Accuracy	No of Attributes/ Indicators Considered
2017	Jhi-Sheng Chou et al[2]	MetaFA-LSSVR	MAPE of 5.628 – 60 days and 6.131 for 90 days	2 Attributes with 7 Indicators
2016	Luckyson Khaidem et al[3]	RF	Short term - 80% Long term – 85%	4 Technical Indicators
2017	Oscar Bustos et al[4]	ANN and SVM	76% and 78%	9 Technical Indicators
2016	Mahajan Shubhrata et al[5]	NB	75.6% – 76.5%	8 Attributes
2014	Xinjie Di[6]	RBF Kernalized SVM	71%	16 Technical Indicators
2014	Bhagwant Chauhan et al[7]	ANN	75%	6 Attributes
2014	Haoming Li et al[8]	Linear, Logistic Regression and SVM	56.65%	5 Attributes, 5 Commodity factors, 2 Foreign Currencies, 1 Interest Rate
2013	Khalid Alkhatib et al[9]	k Nearest Neighbours	67%	3 Attributes
2013	Yuqing Dai et al[10]	SVM, Logistic Regression, QDA, GDA	58.2% - 79.3%	12 Indicators
2012	B.K. Panigrahi et al[11]	RBF	Achieves MAPE of 2.1005 for 30 day prediction	7 Technical Indicators, 5 Attributes

IV. CONSTRAINTS AND FUTURE DIRECTIONS

From the overview of the prediction algorithms for stock the following constraints are identified

In most cases, the algorithms are written specific to the stock market and not in a generic manner. All the algorithms work for a selected national stock market and they predict only companies of those markets and not those of the other markets around the world.

- Even in a particular stock market these algorithms predict effectively only a few listed companies but not all those present in that market.
- Also the algorithms predict effectively only the long term investments which are mainly done by professional traders or organization and not any proper way to support individual traders who generally do short term investments.
- Also the prediction accuracy in most algorithms doesn't go beyond 75% which needs to be optimized or enhanced.

Novel Algorithms may evolve in near future to address the above identified constraints and better algorithms can be used to have a generic approach for stock market prediction.

V. CONCLUSION

Stock market is one of areas where people invest their capital. By predicting the variations in stock prices, proper investment may be done which reduce the risk of their investment. In this work, an overview of the various algorithms that are being used for stock prediction is presented. This work has compared the performance of the various algorithms in the context of the factors used to predict the price variation and their effectiveness. From the overview, we have identified a few alarming constraints which can be rectified by new algorithms and better models

REFERENCES

[1] The Financial Express (2017) "*Why investors lose money in stock market: 10 reasons that will keep you profitable*" [Online]. Available: http://www.financialexpress.com/money/why-investors-lose-money-in-stock-market-how-to-profit-from-stock-market-10-reasons-that-will-keep-you-profitable/879985/

[2] Jui-Sheng Chou, Thi-Kha Nguyen, *"Forward Forecast of Stock Price Using Sliding window Metahuristic-optimized Machine Learning Regression"*, IEEE Transactions on Industrial Infomatics, Vol 99, 2017.

[3] Luckyson Khaidem, Snehanshu Saha, Sudeepa roy Dey, *"Predicting the direction of stock market prices using random forest"*, Applied Mathematical Finance, Volume 23, Issue 2, June 2016.

[4] Oscar Bustos, Alexandra Pomares, Enrique Gonzalez, *"A comparison between SVM and Multilayer*

 DOI : 10.18231/2454-9150.2018.0109

Perceptron in predicting an emerging financial market: Colombian stock market", Congreso Internacional de Innovacion y Tendencias en Ingenieria (CONIITI) sponsored by IEEE, 2017.

[5] Mahajan Shubhrata.D, Deshmukh Kaveri.V, Thite Pranit.R, Samel Bhavana.Y, Chate.P.J, *"Stock Market Prediction and Analysis Using Naïve Bayes"*, International Journal on Recent and Innovation Trends in Computing and Communication,Vol 4, Issue 11,2016.

[6] Xinjie Di, *"Stock trend Prediction with Technical Indicators using SVM"*, Stanford University, 2014.

[7] Bhagwant Chauhan, Umesh Bidave, Ajit Gangathade, Sachin Kale, *"Stock Market Prediction Using Artificial Neural Networks"*, International Journal of Computer Science and Information Technologies, Vol.5(1), 2014.

[8] Haoming Li, Zhijun Yang, Tianlun Li, *"Algorithmic Trading Strategy Based in Massive Data Mining"*, Stanford University, 2014.

[9] Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi, Mohammed K.Ali Shatnawi, *"Stock Price Prediction using K-Nearest Neighbor (kNN) Algorithm"*, International Journal of Business, Humanities and Technology, Volume 3 Issue 3 March 2013.

[10] Yuqing Dai, Yuning Zhang, *"Machine Learning in Stock Price Trend Forecasting"*, Stanford University, 2013.

[11] B.K. Panigrahi, Babita Majhi, Usha Manasi Mohapatra, Rosalin Mahapatra, *"Stock Indices Prediction Using Radial Basis Function Neural Network"*, Swarm, Evolutionary and Memetic Computing, Third International Conference, SEMCCO, 2012.

 DOI : 10.18231/2454-9150.2018.0109